BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 2: e28440
doi: 10.3897/biss.2.28440

OPEN ACCESS

Conference Abstract

Global Genomic Resources for Biodiversity Research

Jonathan Coddington ‡

‡ Smithsonian Institution, Washington, United States of America

Corresponding author: Jonathan Coddington (coddington@si.edu)

Received: 17 Jul 2018 | Published: 19 Jul 2018

Citation: Coddington J (2018) Global Genomic Resources for Biodiversity Research. Biodiversity Information Science and Standards 2: e28440. https://doi.org/10.3897/biss.2.28440

Abstract

Genomic science is revolutionizing and accelerating biodiversity research. For collections-based institutions to continue to lead and support biodiversity research, they must adapt to this new reality. Simultaneously, “big data” is accumulating so rapidly that we have unprecedented capacity to plan strategically to use genomics to advance basic and applied science on multiple fronts. For example, seven “big data” sources (GBIF, ~1B records; BHL, ~3.6M records; NCBI, ~220M records; OToL, 1.9M records; BOLD, ~6.3M records, EOL, ~99K records, and GGBN, ~2M records) collectively offer more than 1.2B records on biodiversity. At the scale of species (~2M described, multiple millions undescribed), these data are still too sparse to permit comprehensive conclusions. At the scale of families (i.e. deeper clades of life), the situation is far more promising: about 9,911 families are known, and relatively few are discovered each year. This suggests that at the family rank (and above), our knowledge of life on Earth is reasonably complete. Approximately 160,000 genera are known, but certainly many new genera await discovery and description, although fewer than new species, and more than new families. Genomics is the fastest way to “bin” species into more inclusive lineages such as genera and families, and is certainly faster than traditional alpha taxonomy. Synergistically, these “big data” answer four important questions at deeper clade levels: What is it? Where is it? What do we know about it? What do we know about its genome? The converse of what we know is what we do not know, another meaning of “dark taxa.” We can use the distribution and density of big data at deeper clade levels (families, genera) quantitatively to analyze “dark taxa,” and

This is an open access article distributed under the terms of the CC0 Public Domain Dedication.

therefore to optimize strategically knowledge and preservation of biodiversity at a global scale. Technicalities of the quantitative prioritization scheme are debatable, but some initial, simple scoring systems can help to prioritize lineages for collection and genetic research so as to most efficiently “illuminate” regions in the tree of life that that are neither preserved, imaged, geolocated, studied, nor known genomically. This analysis presents criteria and goals for collaborating to build a global genomic collection to maximize efficient acquisition of biodiversity genomic knowledge.

Keywords

global genome, taxonomic diversity, biology, gene sequencing

Presenting author

Jonathan Coddington, Global Genome Initiative, Smithsonian Institution